Fast Bayesian Analysis of Spatial Dynamic Factor Models for Large Space Time Data Sets

C. M. Strickland,* D. P. Simpson, I. W. Turner,
R. Denham, K. L. Mengersen

April 23, 2009

Abstract

Remoting sensing is one example where data sets that vary across space and time have become so large that 'standard' approaches employed by statistical modellers for applied analysis are no longer feasible. In this paper, we present a Bayesian methodology, which makes use of recently developed algorithms in applied mathematics, for the analysis of large space time data sets. In particular, a Markov chain Monte Carlo algorithm is proposed for the efficient estimation of spatial dynamic factor models (DFMs). The spatial DFM is specified whereby spatial dependence is modelled though the columns of the factor loadings matrix using a Gaussian Markov random field. Krylov subspace methods are used to take advantage of the sparse matrix structures that are inherent in the model. The methodology is used to analyse remotely sensed data from the Moderate Imaging Spectroradiometer satellite. In particular, the proposed methodology is used in conjunction with high resolution imagery for the classification, in terms of land type, of two regions located in central Queensland, Australia.

Key Words: Bayesian analysis; Markov chain Monte Carlo; MODIS; Gaussian Markov random field; Krylov subspace method; Spatial dynamic factor model

*Corresponding Author: School of Mathematics, GPO Box 2434, Queensland University of Technology, Queensland, 4001, Australia. Email: christopher.strickland@qut.edu.au. Phone: 61 7 3138 8313. Fax: 61 7 3138 2310.

1 Introduction

Land cover and land use maps are often produced using remotely sensed imagery. The imagery is converted to a map using statistical classification approaches, manual interpretation or more commonly a combination of both. When multi-temporal imagery is available the temporal information may be used to either help refine the map, see for example Brown de Colstoun *et al.* (2003), Sugurmaran, Zerr and Prato (2001) and Waske & Braun (2009), or to identify areas of land cover/land use change, see for example Lu *et al.* (2003), Coppin *et al.* (2004) and Department of Natural Resources and Water (DNR) (2008). In the remote sensing literature, multi-temporal imagery typically refers to a series of 2 to 20 images. With such short series of imagery, there is a risk that insufficient variation in the temporal signature is captured, which reduces the ability to distinguish the difference between change in the land cover or land use and inherent variation at the site. More recent satellites deliver much more frequent overpasses, with some products from the Moderate Imaging Spectroradiometer (MODIS) service available daily. Data at this resolution has the potential to provide a far more detailed picture of the landscape, across time and space, but are not suited to techniques that are traditionally used in multi-temporal image analysis. This paper presents a fast Bayesian approach for spatial dynamic factor models (DFM)s, which can be used in the analysis of large space-time data sets. We motivate the approach by considering both land cover and land use applications and demonstrate how the proposed method can be used to assist in the mapping of landscape features that are not easily visible using single date imagery.

DFMs provide a useful representation for high dimensional multivariate time series. One of the main difficulties faced in the analysis of such data is that the number of parameters increase quadratically with the the number of time series. The appeal of the DFM is that it provides a natural specification for dimension reduction. For this reason, DFMs have received much attention in the recent literature; see for example Forni *et al.* (2000), Ritter and Muñoz-Carpena (2006), Sáfadi and Peña (2008) and the references therein.

Lopes, Salazar and Gammerman (2008) extend the standard DFM for spatially correlated data. In particular, the spatial DFM specified by Lopes *et al.* captures spatial dependence through the columns of the factor loadings matrix using a distance based Gaussian random field. A popular alternative, which is utilised in this paper, to the distance based Gaussian random field is to model spatial correlation using a Gaussian Markov random field (GMRF); see Cressie (1993) for details. This modification to the approach of Lopes *et al.* is particularly appealing given the applications considered in the paper. In particular, it allows the use of computationally efficient algorithms that

take advantage of the sparse structures imposed by a GMRF specification, which make an otherwise impractical approach feasible.

The most commonly used approach for sampling from a GMRF is that of Rue (2001). Given a sparse, symmetric positive definite precision matrix $\overline{\mathbf{V}}$, Rue's method proceeds to calculate the Cholesky decomposition $\overline{\mathbf{V}} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower triangular matrix. A sample from a zero mean GMRF with precision matrix $\overline{\mathbf{V}}$ can then be calculated by solving the upper triangular system $\mathbf{L}^T\mathbf{x} = \mathbf{z}$, where $\mathbf{z}$ is a vector of i.i.d. standard normal random variables. When solving large linear systems arising from problems in applied mathematics it is often found that sparse direct methods, such as the Cholesky decomposition, become less efficient as the size of the problem increases. This has encouraged the development of iterative solvers for sparse linear systems and, in particular, Krylov subspace based methods; see for example Saad (2003). To this end, Simpson *et al.* (2008) have developed a Krylov subspace method for sampling from zero mean GMRFs. This approach will be exploited in the proposed algorithm.

The structure of the paper is as follows. In Section 2 the spatial DFM is described, including the state space form (SSF), identification restrictions, and the form of the spatial dependence. In Section 3 the Bayesian Markov chain Monte Carlo (MCMC) methodology that is used for estimation is introduced. In particular, the joint posterior is defined, a Gibbs based sampling scheme is outlined and the algorithms to draw from each block in the Gibbs based scheme are described. In Section 4 the proposed spatial DFM is used in the analysis of remotely sensed data from the MODIS satellite imagery. MODIS provides frequent satellite based reflectance images of the earth's surface that are used extensively in monitoring the condition of the land use and land cover across the globe. Derived products, such as the normalised difference vegetation index (NDVI), which are available in the MOD13Q1, are used by the Department of Natural Resources and Water to analyse cropping and land use patterns and to assess condition and change in natural communities. The proposed method is used in conjunction with Landsat Thematic Mapper (TM) high resolution imagery to classify, in terms of land use, two regions located in central Queensland, Australia. The algorithm proves to be be both fast and scalable, making if feasible to analyse reasonable sized regions of remotely sensed data. Conclusions are drawn in Section 5.

2 Spatial Dynamic Factor Model

The observation equation for the $(p\times 1)$ vector of observations $\mathbf{y}_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,p})^T$, is given, for $t = 1, 2, \ldots, n$, by

$$\mathbf{y}_t = \mathbf{B}_\mu \boldsymbol{\mu}_t + \mathbf{B}_\psi \boldsymbol{\psi_t} + \boldsymbol{\varepsilon}_t; \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \tag{1}$$

where $\mathbf{B}_\mu$ is a $(p \times k_\mu)$ matrix of factor loadings for the $(k_\mu \times 1)$ vector of common trends $\boldsymbol{\mu}_t = (\mu_{t,1}, \mu_{t,2}, \ldots, \mu_{t,k_\mu})^T$ and $\mathbf{B}_\psi$ is a $(p \times k_\psi)$ matrix of factor loadings for the $(k_\psi \times 1)$ vector of common cycles $\boldsymbol{\psi}_t = (\psi_{t,1}, \psi_{t,2}, \ldots, \psi_{t,k_\psi})^T$. The $(p \times 1)$ vector $\boldsymbol{\varepsilon}_t$ is both contemporaneously and serially uncorrelated, with a mean vector of $\mathbf{0}$ and an unknown $(p \times p)$ covariance matrix $\boldsymbol{\Sigma} = \text{diag}\left(\sigma^2_{\varepsilon(1,1)}, \sigma^2_{\varepsilon(2,2)}, \ldots, \sigma^2_{\varepsilon(p,p)}\right)$. Define the block matrix $\mathbf{B} = \left[\begin{array}{cc} \mathbf{B}_\mu & \mathbf{B}_\psi \end{array} \right]$, where $\mathbf{B}$ is a $(p \times k)$ matrix, and $\mathbf{f}_t = \left[\begin{array}{c} \boldsymbol{\mu_t} \\ \boldsymbol{\psi}_t \end{array} \right]$, where $\mathbf{f}_t$ is a $(k \times 1)$ vector, with $k = k_\mu + k_\psi$.

The temporal transition, for $t = 1, 2, \ldots, n-1$, for the vector of common trends $\boldsymbol{\mu}_t$ is defined to follow a multivariate random walk, such that

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \boldsymbol{\zeta}_t; \quad \boldsymbol{\zeta}_t \sim N\left(\mathbf{0}, \mathbf{I}_{k_\mu}\right), \tag{2}$$

where $\mathbf{I}_s$ generically denotes an $(s \times s)$ identity matrix . The i^{th} stochastic cycle, $\psi_{t,i}$ is defined by

$$\left[\begin{array}{c} \psi_{t+1,i} \\ \psi^*_{t+1,i} \end{array} \right] = \mathbf{T}_{\psi(i)} \left[\begin{array}{c} \psi_{t,i} \\ \psi^*_{t,i} \end{array} \right] + \left[\begin{array}{c} \zeta_{t,i} \\ \zeta^*_{t,i} \end{array} \right]; \quad \left[\begin{array}{c} \xi_{t,i} \\ \xi^*_{t,i} \end{array} \right] \sim N\left(\mathbf{0}, \mathbf{I}_2\right), \tag{3}$$

for $i = 1, 2, \ldots, k_\psi$, where $\psi^*_{t,i}$ is an auxiliary variable and $\mathbf{T}_{\psi(i)} = \rho_i \left[\begin{array}{cc} \cos\lambda_i & \sin\lambda_i \\ -\sin\lambda_i & \cos\lambda_i \end{array} \right]$. It is assumed for $i = 1, 2, \ldots, k_\psi$ that ρ_i and λ_i are unknown and that $|\rho_i| < 1$. For notational convenience, define $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{k_\psi})$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{k_\psi})$ and $\boldsymbol{\theta} = \left[\begin{array}{c} \boldsymbol{\rho} \\ \boldsymbol{\lambda} \end{array} \right]$.

The model is completed via the specification of $\boldsymbol{\mu}_1$ and $\boldsymbol{\psi}_1$, whereby it is assumed that

$$\boldsymbol{\mu}_1 \sim N(\mathbf{a}_\mu, \mathbf{P}_\mu) \tag{4}$$

and

$$\boldsymbol{\psi}_1 \sim N(0, \mathbf{P}_\psi). \tag{5}$$

Further, it is assumed that both $\mathbf{a}_\mu$ and $\mathbf{P}_\mu$ are known, and that $\mathbf{P}_\psi = \text{diag}\left((1-\rho_1^2)^{-1}\mathbf{I}_2, (1-\rho_2^2)^{-1}\mathbf{I}_2, \ldots, (1-\rho_{k_\rho}^2)^{-1}\mathbf{I}_2\right)$. For convenience throughout, we define $\mathbf{y} = \left(\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_n^T\right)^T$ and $\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \ldots, \boldsymbol{\alpha}_n^T\right)^T$.

2.1 State Space Form

The SSF is a useful representation for structural time series models as it allows for the straightforward application of generic algorithms that can be used for estimation, such as the Kalman filter. The measurement equation for the DFM in (1), (2) and (3), when expressed in SSF, is given by

$$\begin{aligned} \mathbf{y}_t &= \mathbf{BG}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \\ &= \mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{6}$$

for $t = 1, 2, \ldots, n$, where the $(p \times m)$ matrix $\mathbf{Z} = \mathbf{BG}$ and the $(m \times 1)$ state vector $\boldsymbol{\alpha}_t = \left(\boldsymbol{\mu}^T, \psi_1, \psi_1^*, \psi_2, \psi_2^*, \ldots, \psi_{k_\psi}, \psi_{k_\psi}^*\right)^T$, with $m = k_\mu + 2k_\psi$. The $(k \times m)$ matrix $\mathbf{G}$ is defined such that $\mathbf{G}\boldsymbol{\alpha}_t = \mathbf{f}_t$. The transition equation for the DFM in SSF is defined as

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t; \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{I}_m), \tag{7}$$

for $t = 1, 2, \ldots, n-1$, where $\mathbf{T} = \text{diag}\left(\mathbf{I}_{k_\mu}, \mathbf{T}_{\psi,1}, \mathbf{T}_{\psi,2}, \ldots, \mathbf{T}_{\psi,k_\psi}\right)$. The initial state corresponds to the prior assumptions given in (4) and (5), such that

$$\boldsymbol{\alpha}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1), \tag{8}$$

where $\mathbf{a}_1 = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \mathbf{0} \end{bmatrix}$ and $\mathbf{P}_1 = \text{diag}(\mathbf{P}_\mu, \mathbf{P}_\psi)$.

2.2 Identifiability Restrictions

It is well known that the likelihood for the dynamic factor model in (1), (2) and (3) is not uniquely defined; see for example Harvey (1989). Consequently, it is necessary to impose additional restrictions on the model to ensure that the posterior is identified. Following Harvey, we impose the following structure on the factor loadings matrix that ensures the posterior is identified,

$$\mathbf{B} = \begin{bmatrix} b_{1,1} & 0 & \cdots & 0 \\ b_{2,1} & b_{2,2} & & \vdots \\ b_{3,1} & b_{3,2} & \ddots & 0 \\ \vdots & & & b_{k,k} \\ b_{k+1,1} & \cdots & & b_{k+1,k} \\ \vdots & & & \vdots \\ b_{p,1} & \cdots & \cdots & b_{p,k} \end{bmatrix}. \tag{9}$$

2.3 Spatial Dependence

Following Lopes *et al.* (2008), spatial dependence is modelled through the columns of the factor loading matrix, $\mathbf{B}$. Define $\mathbf{b}_j$ as the j^{th} column of $\mathbf{B}$ and define $\mathbf{R}_j$ as a selection matrix that is defined such that $\mathbf{R}_j\mathbf{b}_j$ is the non-zero component of the j^{th} column of $\mathbf{B}$. Denote $\mathbf{b}_j^* = \mathbf{R}_j\mathbf{b}_j$ and define the prior for $\mathbf{b}_j^*$, for $j = 1, 2, \ldots, k$, as

$$p(\mathbf{b}_j^*) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{b}_j^* - \boldsymbol{\mu}_j\right)^T \mathbf{V}_j \left(\mathbf{b}_j^* - \boldsymbol{\mu}_j\right)\right\}, \tag{10}$$

where $\boldsymbol{\mu}_j$ is the mean of $\mathbf{b}_j^*$ and $\mathbf{V}_j$ is the precision. The specific form of $\boldsymbol{\mu}_j$ and $\mathbf{V}_j$ defines the form of spatial dependence. One possibility, which is pursed by Lopes *et al.* (2008), is to specify (10) such that the spatial dependence is modelled using a Gaussian random field. An alternative approach, which is pursued in this paper, is to model spatial dependence using a GMRF. The advantage of this approach, for the problems addressed in this paper in particular, is that the precision matrix, $\mathbf{V}_j$, for a GMRF is typically sparse. Thus, for large p, computationally efficient samplers can be designed using algorithms from the numerical linear algebra literature. For the analysis in Section 4, we use a zero-mean, first order intrinsic GMRF. In this case, the GMRF in (10) is specified as follows,

$$\begin{aligned}
\boldsymbol{\mu}_j &= 0, \\
\mathbf{V}_j &= \tau_j^{-2}\mathbf{R}_j\mathbf{Q}\mathbf{R}_j^T, \\
Q_{i,j} &= \begin{cases} n_i & i = j \\ -1 & i \sim j \\ 0 & \text{otherwise} \end{cases},
\end{aligned}$$

where $i \sim j$ denotes that region i is adjacent to region j; see Rue and Held (2005) for further details. For notational convenience, denote $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)$.

3 Bayesian Estimation

Bayesian inference about the joint posterior distribution of the dynamic factor model in (1), (2) and (3) is conducted using a hybrid Gibbs/Metropolis-Hastings (MH) MCMC sampling scheme. The joint posterior for the full set of unknown parameters is given by

$$p(\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\theta}, \boldsymbol{\tau}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon) \times p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \times p(\mathbf{B}|\boldsymbol{\tau}) \times p(\boldsymbol{\Sigma}_\varepsilon) \times p(\boldsymbol{\theta}) \times p(\boldsymbol{\tau}), \tag{11}$$

where $p(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon)$ denotes the joint probability density function (pdf) of $\mathbf{y}$ conditional on $\boldsymbol{\alpha}$, $\mathbf{B}$ and $\boldsymbol{\Sigma}_\varepsilon$, $p(\boldsymbol{\alpha}|\boldsymbol{\theta})$ denotes the joint pdf for $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\theta}$, $p(\mathbf{B}|\boldsymbol{\tau})$ denotes the joint pdf for $\mathbf{B}$ conditional on $\boldsymbol{\tau}$ and $p(\boldsymbol{\Sigma}_\varepsilon)$, $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\tau})$ denote the prior pdfs for $\boldsymbol{\Sigma}_\varepsilon$, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$, respectively. From (6) it is clear that

$$p(\mathbf{y}|\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}_\varepsilon|^{-1/2} \exp\{-\frac{1}{2}\sum_{t=1}^{n}(\mathbf{y}_t - \mathbf{Z}\boldsymbol{\alpha}_t)'\boldsymbol{\Sigma}_\varepsilon^{-1}(\mathbf{y}_t - \mathbf{Z}\boldsymbol{\alpha}_t)\},$$

for $t = 1, 2, \ldots, n$. From (7) it follows that

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \propto p(\boldsymbol{\alpha}_1|\boldsymbol{\theta}) \prod_{t=1}^{n-1} p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\theta}), \tag{12}$$

where

$$p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \theta) \propto \exp\{-\frac{1}{2}(\boldsymbol{\alpha}_{t+1} - \mathbf{T}\boldsymbol{\alpha}_t)^T(\boldsymbol{\alpha}_{t+1} - \mathbf{T}\boldsymbol{\alpha}_t)\},$$

for $t = 1, 2, \ldots, n-1$. The prior pdfs for $\mathbf{B}$, $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\theta}$ are assumed to be *a priori* independent.

3.1 Gibbs Scheme

To estimate the joint posterior in (11) a generic Gibbs sampling scheme is defined at iteration j as follows.

1. Sample $\mathbf{B}^{(j)}$ from $p\left(\mathbf{B}|\mathbf{y}, \boldsymbol{\alpha}^{(j-1)}, \boldsymbol{\Sigma}_\varepsilon^{(j-1)}, \boldsymbol{\theta}^{(j-1)}, \boldsymbol{\tau}^{(j-1)}\right)$.
2. Sample $\boldsymbol{\tau}^{(j)}$ from $p\left(\boldsymbol{\tau}|\mathbf{y}, \boldsymbol{\alpha}^{(j-1)}, \mathbf{B}^{(j)}, \boldsymbol{\Sigma}_\varepsilon{}^{(j-1)}, \boldsymbol{\theta}^{(j-1)}\right)$.
3. Sample $\boldsymbol{\alpha}^{(j)}$ from $p\left(\boldsymbol{\alpha}|\mathbf{y}, \mathbf{B}^{(j)}, \boldsymbol{\Sigma}_\varepsilon^{(j-1)}, \boldsymbol{\theta}^{(j-1)}, \boldsymbol{\tau}^{(j)}\right)$.
4. Sample $\boldsymbol{\Sigma}_\varepsilon^{(j)}$ from $p\left(\boldsymbol{\Sigma}_\varepsilon|\mathbf{y}, \boldsymbol{\alpha}^{(j)}, \mathbf{B}^{(j)}, \boldsymbol{\theta}^{(j-1)}, \boldsymbol{\tau}^{(j)}\right)$.
5. Sample $\boldsymbol{\theta}^{(j)}$ from $p\left(\boldsymbol{\theta}, |\mathbf{y}^{(j)}, \boldsymbol{\alpha}^{(j)}, \mathbf{B}^{(j)}, \boldsymbol{\Sigma}_\varepsilon^{(j)}, \boldsymbol{\tau}^{(j)}\right)$.

The following subsections describe the algorithms used to sample from each of the full conditional posteriors of interest in the Gibbs scheme above.

3.2 Sampling B

Defining the $(pk - nr \times pk)$ block diagonal matrix $\mathbf{R} = \text{diag}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_k)$ and the $(pk - nr \times 1)$ vector $\boldsymbol{\beta} = \mathbf{R}\left(\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_k^T\right)^T$, it follows from (10) that the full conditional posterior distribution for $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\right)^T \overline{\mathbf{V}}\left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\right)\right\}, \tag{13}$$

with

$$\overline{\mathbf{V}} = \left(\underline{\mathbf{V}} + \mathbf{R}\left[\left(\sum_{t=1}^{n} \mathbf{f}_t \mathbf{f}_t^T\right) \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}\right]\mathbf{R}^T\right)$$

and

$$\begin{aligned}\bar{\boldsymbol{\beta}} &= \overline{\mathbf{V}}^{-1}\left(\underline{\mathbf{V}}\boldsymbol{\mu} + \mathbf{R}\left[\sum_{t=1}^{n} \mathbf{f}_t \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}\right]\mathbf{y}_t\right) \\ &= \overline{\mathbf{V}}^{-1}\boldsymbol{\omega},\end{aligned}$$

where $\boldsymbol{\omega} = \left(\underline{\mathbf{V}}\boldsymbol{\mu} + \mathbf{R}\left[\sum_{t=1}^{n} \mathbf{f}_t \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}\right]\mathbf{y}_t\right)$, $\underline{\mathbf{V}} = \text{diag}(\mathbf{V}_1\mathbf{V}_2, \ldots, \mathbf{V}_k)$, $\boldsymbol{\mu} = \left(\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \ldots, \boldsymbol{\mu}_k^T\right)^T$ and $\otimes$ is the Kronecker product.

Given (13), it is straightforward to draw from $\boldsymbol{\beta}$, and equivalently $\mathbf{B}$, using the Cholesky decomposition. However, as $\overline{\mathbf{V}}$ is sparse, this approach needs to be carefully implemented to avoid inefficiency. The key to maintaining efficiency is to employ a graph reordering scheme to reduce the bandwidth of the reordered precision matrix. This reduces the number of required operations from $\mathcal{O}\left((pk - nr)^3\right)$ to $\mathcal{O}\left((pk - nr)\, b^2\right)$, where b is the bandwidth of the reordered matrix. For precision matrices arising from two dimensional second order random walks, it can be shown that the required number of operations is $\mathcal{O}((pk - nr)^{3/2})$; see Golub and van Loan (1996) and Rue (2001) for further details.

While direct methods for solving sparse linear systems have well known operation counts, it is often found that iterative methods for large sparse systems can perform better on some practical problems. Therefore, for the remainder of this section, we will outline an iterative method for sampling from (13) that is introduced in Simpson *et al.* (2008). It is interesting to note that any decomposition of the form $\overline{\mathbf{V}} = \mathbf{C}\mathbf{C}^T$ can be used to sample from a GMRF. In particular, the choice $\mathbf{C} = \overline{\mathbf{V}}^{1/2}$ leads to an iterative method based on Krylov subspaces. With this choice of $\mathbf{C}$, the procedure for sampling from $\boldsymbol{\beta}$ is achieved using the following algorithm.

1. Solve $\overline{\mathbf{V}}\bar{\boldsymbol{\beta}} = \boldsymbol{\omega}$ using the conjugate gradient method; see Golub and van Loan, (1996).

2. Draw $((pk - nr) \times 1)$ vector $\mathbf{z}$ where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_{pk-nr})$.

3. Approximate $\mathbf{w} = \overline{\mathbf{V}}^{-1/2}\mathbf{z}$, using the Lanczos algorithm; see below.

4. Take $\tilde{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} + \mathbf{w}$ as a draw from (13).

The most common method for solving large sparse symmetric positive definite linear systems, such as the one in Step 1, is the conjugate gradient method. Although not always framed as such, the conjugate gradient method is a member of the Krylov subspace based family of numerical methods. The J–dimensional Krylov subspace generated by a matrix $\overline{\mathbf{V}}$ and a vector $\mathbf{z}$ is defined as

$$\mathcal{K}_J(\mathbf{z}, \overline{\mathbf{V}}) = \text{span}\{\mathbf{z}, \overline{\mathbf{V}}\mathbf{z}, \overline{\mathbf{V}}^2\mathbf{z}, \ldots, \overline{\mathbf{V}}^{J-1}\mathbf{z}\}.$$

It follows that Krylov subspace methods *do not* require the full matrix $\overline{\mathbf{V}}$, but only the ability to form the matrix–vector product $\overline{\mathbf{V}}\mathbf{z}$. This can be a particularly efficient way of exploiting the structure of $\overline{\mathbf{V}}$ and is a major advantage of the Krylov subspace based method over the Cholesky method.

A Krylov subspace method can also be used to calculate $\mathbf{w} = \overline{\mathbf{V}}^{-1/2}\mathbf{z}$. The first step of this method is to build an orthonormal basis for the Krylov subspace generated by $\overline{\mathbf{V}}$ and $\mathbf{z}$. This is achieved using the Lanczos algorithm, which is a clever rearrangement of the familiar Gram–Schmidt algorithm; see for example Stewart (2001) for further details. The Lanczos algorithm for $j = 1, 2, \ldots, J$ can be summarised as follows,

$$\begin{aligned} \mathbf{v}_j &= \overline{\mathbf{V}}\mathbf{u}_j, \\ a_j &= \mathbf{u}_j^T\mathbf{v}_j, \\ \mathbf{v}_{j+1} &= \mathbf{v}_j - a_j\mathbf{u}_j - d_{j-1}\mathbf{u}_{j-1}, \\ d_j &= \|\mathbf{v}_{j+1}\|, \\ \mathbf{u}_{j+1} &= \mathbf{v}_{j+1}/d_j, \end{aligned} \tag{14}$$

with $\mathbf{u}_1 = \frac{\mathbf{z}}{\|\mathbf{z}\|}$ and $d_0 = 0$. The algorithm in (14) generates the Lanczos relation,

$$\overline{\mathbf{V}}\mathbf{U}_J = \mathbf{U}_{J+1}\overline{\mathbf{T}}_J = \mathbf{U}_J\mathbf{T}_J + d_J\mathbf{u}_{J+1}\mathbf{e}_J^T, \tag{15}$$

where

$$\mathbf{T}_J = \begin{bmatrix} a_1 & d_1 & & & \\ d_1 & a_2 & d_2 & & \\ & d_2 & a_3 & \ddots & \\ & & \ddots & \ddots & d_{J-1} \\ & & & d_{J-1} & a_J \end{bmatrix}$$

is a $(J \times J)$ tridiagonal matrix and the columns of $\mathbf{U}_J = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_J)$ form an orthonormal basis for $\mathcal{K}_J(\overline{\mathbf{V}}, \mathbf{z})$ and the $(J \times 1)$ vector $\mathbf{e}_J$ defined such that the J^{th} element is one and the other entries are zero; see Saad (2003) for further details. This basis can be used to calculate the sample $\mathbf{w} = \overline{\mathbf{V}}^{-1/2}\mathbf{z}$ to within a specified accuracy ϵ. The following algorithm, which is described in greater depth in Simpson *et al.* (2008), details the procedure.

1. Apply the Lanczos algorithm in (14) until $l_{min}^{-1/2}\|\mathbf{z}\| \left|\mathbf{e_J}^T \mathbf{T_J}^{-1} \mathbf{e_1}\right| < \epsilon$, where l_{min} is the smallest eigenvalue in $\mathbf{T}_J$.
2. Calculate the diagonalisation $\mathbf{T}_J = \mathbf{P}_J \Lambda_J \mathbf{P}_J^T$.
3. Calculate $\mathbf{w} = \|z\| \mathbf{U}_J \mathbf{P}_J \Lambda_J^{-1/2} \mathbf{P}_J^T \mathbf{e}_1$, where $\Lambda_J^{-1/2} = \text{diag}(l_1^{-1/2}, l_2^{-1/2}, \ldots, l_J^{-1/2})$.

The exit condition, which is used in the first step to adaptively choose the subspace size J, is strongly related to the exit condition that arises when solving linear systems; see Saad for further details.

3.3 Sampling $\boldsymbol{\tau}$

In sampling $\boldsymbol{\tau}$, for $j = 1, 2, \ldots, k$, τ_j is drawn separately from its full conditional posterior distribution. It is assumed that

$$p(\tau_j) \sim IG\left(\frac{\underline{\nu}}{2}, \frac{\underline{S}}{2}\right), \tag{16}$$

where $IG()$ refers to the inverted gamma distribution and both $\underline{\nu}$ and $\underline{S}$ are hyperparameters. It follows that the posterior distribution is given as

$$p\left(\tau_j | \mathbf{y}, \boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\theta}\right) \propto \tau_j^{-(\bar{\nu}+1)} \exp\left\{-\frac{\bar{S}}{2\tau_j}\right\},$$

where $\bar{\nu} = \underline{\nu} + p$ and $\bar{S} = \underline{S} + \mathbf{b}_j^T \mathbf{Q}_j \mathbf{b}_j$.

3.4 Sampling $\boldsymbol{\alpha}$

It is straightforward to sample the state vector, $\boldsymbol{\alpha}$, from its full conditional posterior distribution,

$$p(\boldsymbol{\alpha} | \mathbf{y}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\theta}, \boldsymbol{\tau}), \tag{17}$$

using a simulation smoother. Frühwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002) all provide alternative simulation smoothing algorithms based on the Kalman

filter. Recently, Strickland *et al.* (2008) proposed a simulation smoothing algorithm that is substantially more computationally efficient than the aforementioned algorithms when sampling the state vector from multivariate linear Gaussian state space models. The algorithm proposed by Strickland *et al.* is a modified version of the simulation smoother of Durbin and Koopman whereby the moments of interest are calculated using the filtering and smoothing algorithms of the univariate representation of the state space model; see Anderson and Moore (1979), Koopman and Durbin (2000) and Strickland *et al.* for further details.

The algorithm of Strickland *et al.* samples the state vector in $\mathcal{O}(pm^2n)$ operations compared to the $\mathcal{O}(p^3n)$ operations required by the Kalman filter based approaches. This leads to substantial savings for high dimensional multivariate time series analysis where $p \gg m$. An alternative approach that is implemented in this paper, which has been developed specifically for DFMs by Jungbacker and Koopman (2008), achieves substantial gains over the aforementioned approaches. Jungbacker and Koopman propose a method that uses a transformation to achieve a reduction in the dimension of the statespace that is required for the estimation of the factors. They show that only a k-dimensional multivariate state space model, with m states, is required to draw from (17). As such, only $\mathcal{O}(m^3n)$ operations are required to sample $\boldsymbol{\alpha}$ from its full conditional posterior distribution.

Jungbacker and Koopman (2008) transform the measurement equation in (6) by pre-multiplying it with $\mathbf{A}^L$, where

$$\mathbf{A}^L = \left(\mathbf{B}^T\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right)^{-1}\mathbf{B}\boldsymbol{\Sigma}_\varepsilon^{-1},$$

such that

$$\mathbf{y}_t^L = \mathbf{G}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t^L, \tag{18}$$

with $\mathbf{y}_t^L = \mathbf{A}^L\mathbf{y}_t$ and $\boldsymbol{\varepsilon}_t^L = \mathbf{A}^L\boldsymbol{\varepsilon}_t$. The measurement disturbance $\boldsymbol{\varepsilon}_t^L$ is serially uncorrelated, with a mean vector $\mathbf{0}$ and a covariance matrix $\left(\mathbf{B}^T\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right)^{-1}$. Jungbacker and Koopman show that standard simulation smoothing algorithms can then be run on the state space model in (18), (7) and (8) to obtain a draw from (17).

3.5 Sampling $\boldsymbol{\Sigma}_\varepsilon$

Sampling $\boldsymbol{\Sigma}_\varepsilon$ from its posterior distribution is completed by individually sampling σ_{ii}^2, for $i = 1, 2, \ldots, p$, from

$$p(\sigma_{ii}|\mathbf{y}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}) \;\propto\; p(y_i|\mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta})p(\sigma_{ii}).$$

It assumed *a priori* that

$$\sigma_{ii} \sim IG\left(\frac{\underline{a}}{2}, \frac{\underline{b}}{2}\right), \tag{19}$$

where $\underline{a}$ and $\underline{b}$ are both hyperparameters. It follows that the posterior distribution for σ_{ii} is

$$p(\sigma_{ii}^2|\mathbf{y}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}) \propto \sigma_{ii}^{(\bar{a}+1)} \exp\left\{-\frac{\bar{b}}{2\sigma_{ii}^2},\right\}$$

where $\bar{a} = \underline{a} + n$, $\bar{b} = \underline{b} + (y_{t,i} - \delta_{t,i})^2$ and $\delta_{t,i}$ is the i^{th} component of $\mathbf{Z}\boldsymbol{\alpha}_t$.

3.6 Sampling $\boldsymbol{\theta}$

Given (11), the full conditional posterior distribution for $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\lambda} \end{bmatrix}$ is given by

$$p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\alpha}, \boldsymbol{\tau}) \propto p(\boldsymbol{\alpha}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}), \tag{20}$$

where $p(\boldsymbol{\alpha}|\boldsymbol{\theta})$ is defined in (12) and it is assumed *a priori* that

$$p(\boldsymbol{\theta}) = p(\rho_1) \times p(\rho_2) \times \cdots \times p\left(\rho_{k_\psi}\right) \times p(\lambda_1) \times p(\lambda_2) \times \cdots p\left(\lambda_{k_\psi}\right).$$

In the empirical analysis, the prior for ρ_j, $j = 1, 2, \ldots, k_\psi$, is assumed to follow a beta distribution, such that

$$\rho_j \sim \text{beta}(a_\rho, b_\rho), \tag{21}$$

where a_ρ and b_ρ are hyperparameters. A uniform prior is assumed for λ_j, $j = 1, 2, \ldots, k_\psi$, such that

$$p(\lambda_j) \sim \text{U}(a_\lambda, b_\lambda), \tag{22}$$

where a_λ and b_λ are hyperparameters.

To sample from (20), ρ_j and λ_j are drawn individually from their posterior distributions, for $j = 1, 2, \ldots, k_\psi$, using the random walk MH algorithm; see Robert and Casella (1999) for details.

4 Empirical Analysis

Land cover and land use over Queensland has been mapped extensively using remote sensing data since 1988; see Witte *et al.* (2006) and DNR (2008) for

further details. Landsat TM and Landsat ETM are used extensively as these products provide coverage across the state at a spatial resolution that is sufficient to identify landscape features. The archive used typically incorporates one or two images per year. This multi-temporal imagery is used to assist in the classification of land use or in the identification of land cover change. Classification is generally approached by manual interpretation of the images, while land cover change is either investigated by manual interpretation, see Witte (2006), or by a semi-automated non-parametric classification, see DNR (2008). With irregular, infrequent imagery, it can be difficult to gain an understanding of landscape features that vary over time, and neither the manual interpretation nor the non-parametric classifier are well suited to longer time series. The MODIS products provide far more frequent imagery across Queensland, but at a coarser resolution. In this section, we show how the frequent coarse data can be used in combination with the less frequent higher resolution imagery to gain insight into the landscape features. The MODIS imagery used is the Normalised Difference Vegetation Index (NDVI) from the MODIS MOD13Q1. The MODIS imagery is available at 16 day intervals with a spatial resolution of 250m. This is considered moderate to coarse, so features less than a few hectares in size are generally difficult to identify from any single MODIS image. NDVI is a commonly used index that provides a numerical representation of the amount of live green vegetation. The ability to detect differences in the spatial and temporal distribution of NDVI has many applications in remote sensing; see for example Hansen *et al.* (2002), Huete *et al.* (2002) and Running *et al.* (2004).

Both analyses use the spatial DFM in (1), (2) and (3), and results are based on 10000 iterations of the MCMC sampling scheme, which are stored after an initial burn-in period of 5000 iterations. The hyperparameters $\underline{\nu}$ and $\underline{S}$ in (16) are set to 3.0 and 0.3, respectively, implying a prior mean of 0.44 and a variance of 0.11 for τ_j, for $j = 1, 2, \ldots, k$. The hyperparameters $\underline{a}$ and $\underline{b}$ in (19) are also set to 3.0 and 0.3, respectively, for σ_{ii}, for $i = 1, 2, \ldots, p$. The hyperparameters a_ρ and b_ρ in (21) and a_λ and b_λ in (22) are set to 15, 1.5, $\frac{1.2\pi}{20}$ and $\frac{2.8\pi}{20}$, for $j = 1, 2, \ldots, k_\psi$, which implies prior means of 0.91 and 0.31 and prior variances of 0.01 and 0.01 for ρ_j and λ_j, respectively.

4.1 Theodore data set

The first data set consists of 197 temporal observations of NDVI from the MODIS MOD13Q1 satellite, over a region containing 900 pixels representing an area of 5,625 hectares. Images between February 18, 2000 to September 13, 2008 are used. The study area is in central Queensland, Australia, approximately 70km west of Theodore, and includes two land cover classes;

grazing land and woodland from a national park. The aim of this analysis is to use the factor model to differentiate land use using both the temporal and spatial information that is in the imagery.

Table 1: Estimation results for Theodore data set, using 10000 iterations of the MCMC sampling scheme, with a burn-in of 5000 iterations.

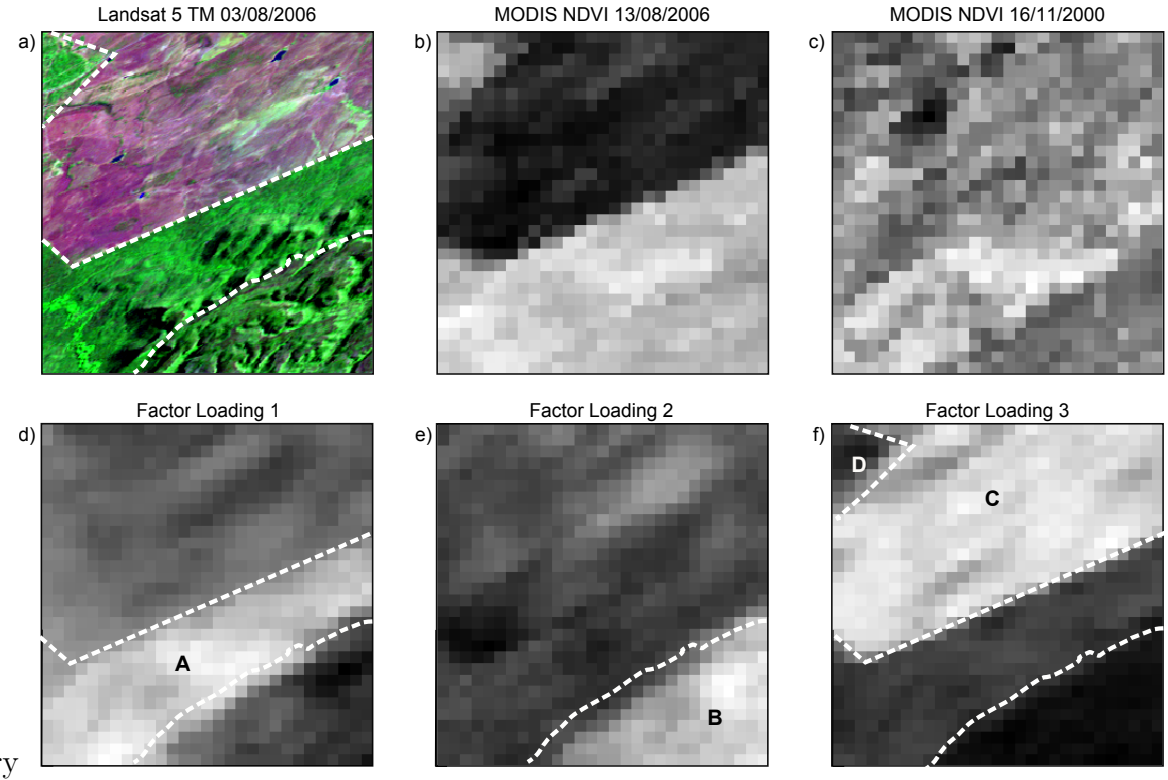
	Model 1 k_μ=1,k_ψ=0	Model 2 k_μ=2,k_ψ=0	Model 3 k_μ=3,k_ψ=0	Model 4 k_μ=4,k_ψ=0	Model 5 k_μ=3,k_ψ=1
BIC	329762	139002	**127676**	159635	423909
Time	217	464	686	919	922

Table 1 contains output from the MCMC estimation of the spatial DFM on the Theodore data set. Five models were considered in the analysis. Models 1, 2, 3 and 4 contain only trend components, whilst Model 5 contains both trend and seasonal components. For each of the models under consideration the Bayesian information criterion (BIC) and the time taken in seconds is reported. A 2.4 GHz Pentium Core 2 Duo with 2 Gigabytes of RAM is used for the computation, though out.

As indicated in Table 1, Model 3 is preferred based on the BIC. Model 3 contains three trend components. The algorithm is clearly very fast as it takes approximately 11 minutes to complete 15000 iterations. This includes the estimation of approximately 3600 parameters, with around 180000 observations across time and space. It is also evident that the algorithm scales almost linearly for the given problem.

Figure 1 (d)-(f) contains a plot of the estimated factor loadings for Model 4. The 30 by 30 region depicted in the plot directly corresponds to the 30 by 30 pixel region of interest in the study area. Figure 1(a) shows a false colour image of the area using bands 5,4 and 2 from a Landsat-TM scene captured on 3 Aug 2008. This scene is chosen from 15 available images between June 2000 and September 2008 since it most clearly shows the landscape features in the region. The MODIS NDVI image from approximately the same date is given in Figure 1(b). At this particular date, a number of features are clearly visible in both the Landsat-TM image and the NDVI image. This is not always so apparent, and Figure 1(c) illustrates this. On this date, differences in NDVI between features, at this site, are far less obvious. The separation into landscape components is clear in the factor loadings shown in Figure 1(d)-(f). Regions marked A and B in factor loadings 1 and 2 represent woodland communities, and the area marked C in factor loading 3 represents grassland. While the distinction between grassland and woodland

Figure 1: Plots of the marginal posterior mean estimates of the factor loadings, for the Theodore data set.



January

Table 2: Estimation results for the cropping data set, using 10000 iterations of the MCMC sampling scheme, with a burn-in of 5000 iterations.

	Model 1 k_μ=1,k_ψ=0	Model 2 k_μ=2,k_ψ=0	Model 3 k_μ=3,k_ψ=0	Model 4 k_μ=4,k_ψ=0	Model 5 k_μ=3,k_ψ=1
BIC	324074	302603	277622	277956	**272437**
Time	224	460	676	882	892

is clear in some of the individual MODIS scenes (Figure 1(b) for example), the subdivision of the woodland is not. From Figure 1(a), however, it is possible to see differences in the vegetation community on the ridge tops in the bottom right hand corner of the image. It appears that the NDVI of this area has a slightly different temporal signature than the lower forested slopes. This subtle difference is difficult to identify, without time consuming manual inspection of imagery, and could easily be missed, if imagery when the distinction is visible is not available. Interpretation of the factor loadings is not, however, always clear. Factor loadings 1 to 3 each highlight a particular landscape component, but the woody patch of vegetation in the top left of the image is not highlighted particularly well, in any of the factors. It presumably behaves slightly differently than any of the other areas and can only be identified after all other areas are interpreted.

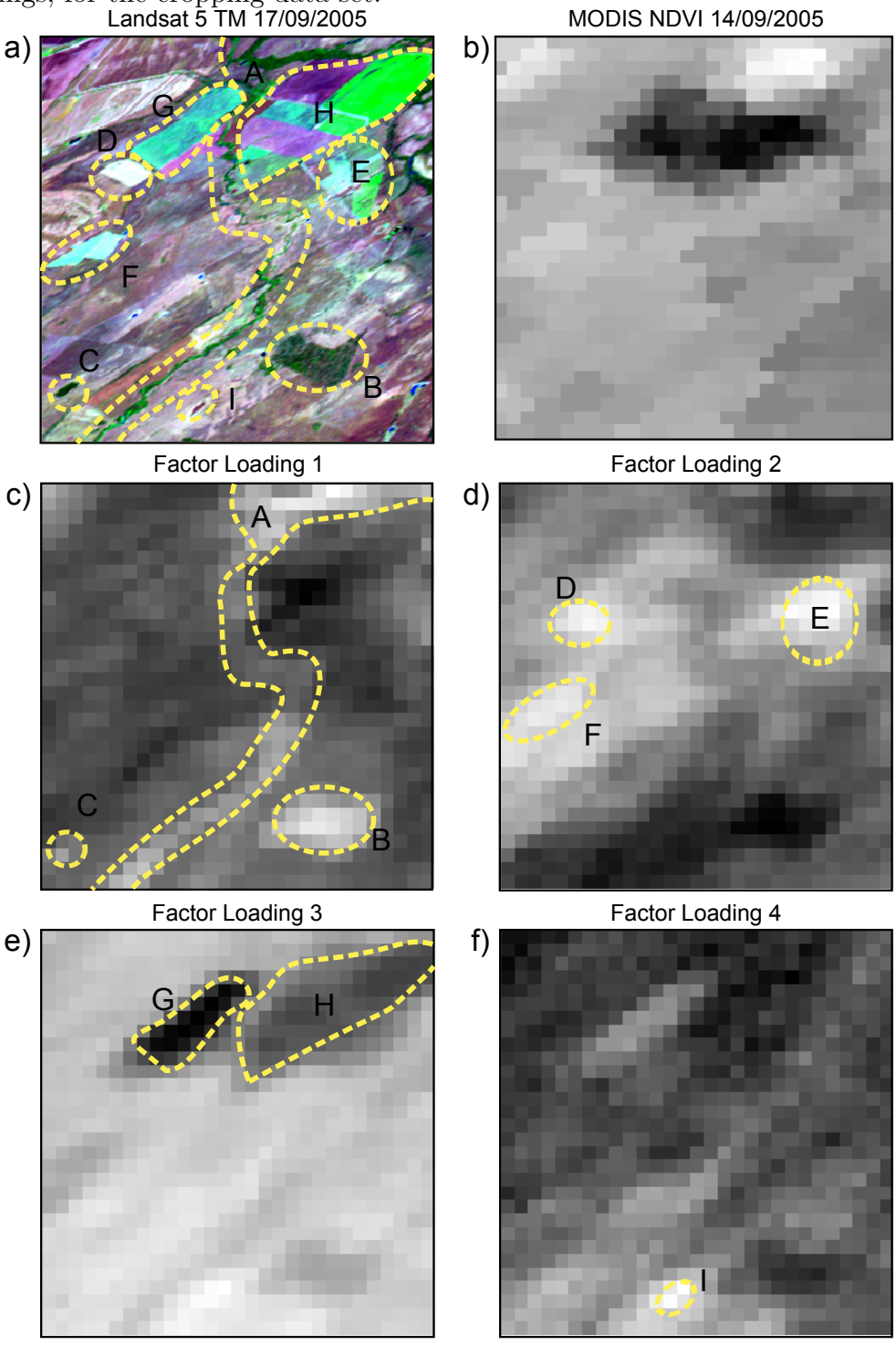
4.2 Cropping data set

The second data set is drawn from an area that has three main land uses; grazing, cropping and woody cover. The area is again a 30 by 30 subset of the MOD13Q1 NDVI product, with observations at 201 dates from February 2000 until November 2008. The site is located approximately 30km north west of Moura, central Queensland, Australia.

Table 2 reports output from the MCMC estimation of the spatial DFM, on the cropping data set. The preferred model, based on the BIC, contains three levels and one cycle. As in the Theodore data set, the computational expense scales almost linearly with respect to the problem size.

Figure 2 summarises the factor loadings. Figure2(a) is a false colour image of bands 5,4 and 2 from Landsat-TM for the region using imagery captured in September 2005. This scene was chosen from 18 available Landsat images as one that most clearly shows the landscape components. From this scene we can see areas of woody vegetation (regions A,B and C) and cropping areas (regions D-H). Areas unlabelled are grazing. The MODIS NDVI scene from

Figure 2: Plots of the marginal posterior mean estimates of the factor loadings, for the cropping data set.



approximately the same date (Figure 2 b)) captures only the most distinct features. In this case, the portion of actively growing crop with high NDVI in area H shows up as an area of light pixels, and areas of fallow with very low NDVI show as patches of dark pixels. Other features are much harder to distinguish. Areas of woody vegetation, for example, do not appear any different to the general area of grazing. In contrast, the plots of the factor loadings (Figure 2 b)-f)) allow us to identify a number of features. Light pixels in factor loading 1 show the distribution of woody areas in the subset. Riverine woodland runs through the centre of the image (region A) and spreads out toward the top. A distinct patch of woodland is shown in the bottom right (region B), with a much smaller patch on the bottom left (region C). Region C is only around 5 hectares in size and would normally be too small to be identified on MODIS imagery. Factor loading 3 shows two distinct dark areas on the top right of the image (regions G and H). These correspond to two areas of crop land. Factor loading 3 shows three areas marked by light pixels (regions D,E and F). Careful inspection of the Landsat imagery reveals that these areas are also cropping areas, but differ from the areas distinguished in factor loading 4 in that they were areas of winter cropping only as opposed to a mix of winter and summer cropping. The spatial DFM thus allows us to identify features of the same land use (cropping) that may be under different management regimes. This management effect is very difficult to determine without manually inspecting imagery over a range of dates and seasons. Cropping/grazing rotations similarly are difficult to distinguish, as areas in the grazing portion of the rotation appear identical to surrounding grazing areas, even in high resolution imagery. Presence of contour banks and paddock shape suggest that the area that is in the top left of the region have at some time been cropped. The dynamic factor model indicates that this region behaves in the same manner as the grazing area in the region, and thus has not been cropped between 2000 and 2008. Manual inspection of the Landsat-TM imagery confirms this.

A final feature, labelled as region I is visible in factor loading 4 (Figure 2 f)). This is an area of woody vegetation that appears to have died some time between July 2003 and February 2004. Again this feature is too small to be identified in any single MODIS image, however the temporal signature for that area provides enough information to distinguish it, and also to identify it as different to other areas of woody vegetation.

5 Conclusions

In this paper, an MCMC methodology for the Bayesian estimation of spatial DFMs is introduced. The spatial DFM of interest is specified so that spatial correlation is captured using a GMRF. The methodology makes use of Krylov subspace methods to exploit the sparse structure of the precision matrix of the GMRF to efficiently sample from the posterior distributions of interest.

The methodology is applied to two remotely sensed MODIS data sets, with each data set being comprised of 900 pixels and around 200 temporal observations. The aim of the analysis is to classify the regions of interest, with respect to land class. The results show the ability to distinguish between subtle differences in land use or land management. In particular, from the model output differences in woodland areas, in the first analysis, and differences in the management of crops, in the second analysis, were apparent where in both cases this difference is not clear when examining any of the individual MODIS images.

High temporal resolution remotely sensed imagery is becoming increasingly common. Such image time series should provide more information on the spatial and temporal characteristics of landscape features, but are not easily incorporated into existing manual or statistical classification approaches. The spatial DFM, presented in this paper, provides an approach that greatly helps remote sensing analysts to characterise landscape features. As we show, the factor loadings are readily interpretable, and can be used to identify features that are not easily discernible with single date imagery. From the analyses, it is clear that the proposed estimation scheme for the spatial DFM is fast, taking only a few minutes to estimate several thousand parameters given approximately 180000 observations across time and space. Further, for the problems considered, the algorithm can also be seen to scale linearly to the size of the problem. The speed and scalability of estimation is of key importance for the method to be useful to analysts of remotely sensed data.

Acknowledgments

This research has been supported by an Australian Research Council Linkage Grant; LP668185. We would like to thank Hedibert Lopes, Birgir Hrafnkelsson and Sylvia Frühwirth-Schnatter for helpful comments and suggestions. The numerical results in the paper are produced using C++ code, except where otherwise stated. The code makes use of the Template Numerical Toolkit, which can be found at http://math.nist.gov/tnt/index.html. Software for the random number generator used can be found at http://www.agner.org

/random/randomc.htm.

References

[1] Anderson, B. D. O and Moore, J. B. (1979) Optimal Filtering. *Prentice Hall*, Englewood Cliffs.

[2] Brown de Colstoun, E. C., Story, M. H., Thompson, C., Commisso, K., Smith, T. G. and Irons, J. R., (2003) National Park vegetation mapping using multi-temporal Landsat 7 data and a decision tree classifier Remote Sensing of Environment, 85, p-316-327.

[3] Cressie, N., (1993). Statistics for spatial data. *Wiley*, New York.

[4] Carter, C. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, 81, 541-553.

[5] Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004) Digital change detection methods in ecosystem monitoring: A review. International Journal of Remote Sensing, 25, 1565−1596.

[6] de Jong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika,* 82, 339-350.

[7] Department of Natural Resources and Water (2008) Land cover change in Queensland 20052006: A Statewide Landcover and Trees Study (SLATS). *Department of Natural Resources and Water*, Brisbane, http://www.nrw.qld.gov.au/slats/report.html.

[8] Durbin, J. and Koopman, S.J. (2002) A simple and efficient simulation smoother for time series models. *Biometrika*, 81, 341-353.

[9] Forni, M., Hallin, M., Lippi, M. and Reichlin, L., (2000) The generalized dynamic factor model: identification and estimation. *Review of Economics and Statistics*, 82, 540-55.

[10] Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15, 183-202.

[11] Golub, G. H. and van Loan, C. F., (1996) Matrix Computations. *John Hopkins University Press*, Baltimore.

[12] Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., Dimiceli, C. and Carroll, M., (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83, 303-319.

[13] Harvey, A. C., (1989) Forecasting structural time series and the Kalman filter. *Cambridge University Press*, Cambridge.

[14] Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indicies. *Remote Sensing of Environment*, 83, 195-213.

[15] Jungbacker, B. and Koopman, S. J., (2008) Likelihood-based analysis for dynamic factor models. *Working Paper*, Tinbergen Institute, Amsterdam.

[16] Koopman, S.J. and Durbin, J. (2000) Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21, 281-96.

[17] Lopes, H. F., Salazar, E. and Gammerman, D., (2008) Spatial dynamic factor analysis. *Bayesian Analysis*, 3, 1-34.

[18] Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004) Change detection techniques. International Journal of Remote Sensing, 25, 2365—2407.

[19] Ritter, A. and Muñoz-Carpena, R., (2006) Dynamic factor modeling of ground and surface water levels in an agricultural area adjacent to Everglades National Park. *Journal of Hydrology*, 317, 340-354.

[20] Robert, C. P. and Casella, G. (1999) Monte Carlo Statistical Methods. *Springer-Verlag*, New York.

[21] Rue, H., (2001) Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B*, 63, 325-338.

[22] Rue, H. and Held, L., (2005) Gaussian Markov random fields. *Chapman and Hall/CWC*, Florida.

[23] Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M. and Hashimoto, H., (2004) A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 54, 547-560.

[24] Saad, Y., (2003) Iterative Methods for Sparse Linear Systems, 2nd Edition. *SIAM*, Philadelphia.

[25] Sáfadi, T. and Peña, D., (2008) Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. *Environmetrics*, 19, 582-601.

[26] Simpson, D. P., Turner, I. W. and Pettitt, A. N., (2008) Fast sampling from a Gaussian Markov random field using Krylov subspace approaches. *QUT Eprints 14376*, Brisbane.

[27] Stewart, G. W., (2001) Matrix Algorithms, Volume II: Eigensystems. *SIAM*, Philadelphia.

[28] Strickland, C. M., Turner, I. W., Denham, R. and Mengersen, K. L., (2008) Efficient Bayesian estimation of multivariate state space models. *QUT Eprints 12499*, Brisbane.

[29] Sugurman, R. Zerr, D. and Prato, T. (2002) Improved urban land cover mapping using multi-temporal IKONOS images for local government planning Canadian Journal of Remote Sensing 28(1) 90-95.

[30] Waske, B., Braun, M., (2009) Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, doi:10.1016/j.isprsjprs.2009.01.003

[31] Witte, C., van den Berg, D., Rowland, T., O'Donnell, T., Denham., R, Pitt, G. and Simpson, J., (2006). Mapping Land Use in Queensland – Technical Report on the 1999 Land Use Map for Queensland. *Department of Natural Resources, Mines and Water*, Brisbane.